JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2000, **74,** 363–375 NUMBER 3 (NOVEMBER)

EXPERIMENTAL ANALYSIS OF HUMAN VOCAL BEHAVIOR: APPLICATIONS OF SPEECH-RECOGNITION TECHNOLOGY

OLIVER WIRTH, PHILIP N. CHASE, AND KEVIN J. MUNSON

WEST VIRGINIA UNIVERSITY

Recent developments in speech recognition make it feasible to apply the technology to study vocal behavior. The present study illustrates the use of this technology to establish functional stimulus classes. Eight students were taught to say nonsense words in the presence of arbitrarily assigned sets of symbols consistent with three three-member experimenter-defined stimulus classes. Computer-controlled speech-recognition software was used to record, analyze, and differentially reinforce vocal responses. When the stimulus classes were established, students were taught to say a new nonsense word in the presence of one member of each stimulus class. Transfer of function was tested subsequently to determine if the novel stimulus names transferred to the remaining stimulus class members. Most subjects required two iterations of the training and testing procedures before transfer occurred. The data illustrate the usefulness of recording vocal behavior during stimulus control procedures and demonstrate the use of speech-recognition technology. The paper also describes the current state of speech-recognition technology and suggests several other areas of research that might benefit from using vocal behavior as its primary datum.

Key words: transfer of function, functional equivalence, speech recognition, naming, vocal behavior, humans

The experimental analysis of human behavior continues to evolve with advancements in research methods and instrumentation. Technological advances, particularly those involving computer hardware and software, have led to improvements in experimental design and data-recording procedures. Speech recognition is one technology that offers the experimental analysis of behavior further sophistication and new avenues for research.

Speech recognition is not a new technology. Considerable interest already has been directed toward its practical uses, especially in the workplace (see Milheim, 1993, for a review). Because this interest has been driven in large part by the promise of improved efficiency in human–computer interactions, much of the interest has come from a human factors or engineering perspective (e.g., Tucker & Jones, 1991). There has been limited application of speech recognition to the study of basic psychological processes or for developing basic theories of learning.

Some basic behavior-analytic research during the 1960s and 1970s, however, was directed toward the analysis of vocal operants (see review by Eshleman, 1991). These early studies most often were directed at schedule control of vocalizations with various species, including humans (Cross & Lane, 1962; Flanagan, Goldiamond, & Azrin, 1958; H. Lane, 1960, 1964; H. Lane & Shinkman, 1963; Miller, 1968; Routh, 1969; Shearn, Sprague, & Rosenzweig, 1961), dogs (Salzinger, Waller, & Jackson, 1962), monkeys (Leander, Milan, Jasper, & Heaton, 1972), cats (Molliver, 1963), and mynah birds (Hake & Mabry, 1979). Typically, these studies relied on the development and use of voice-activated relays to record occurrences of vocal utterances above preset thresholds of amplitude. These devices were fairly accurate in the detection of vocal utterances and could be easily programmed to deliver consequences contingent upon the occurrence, rate, or, at best, the pitch of vocal responding. The early speech-recognition systems, however, were not without weaknesses. These systems were plagued with recognition errors and required expensive computers. Most limiting perhaps was the inability of voice-activated relays to record the content of vocalizations and differentially reinforce various specified topographies (Baron & Journey, 1989).

Baron and Journey (1989) devised a computer-controlled speech-recognition system

Address all correspondence to Oliver Wirth, who is now at the National Institute for Occupational Safety and Health (NIOSH), 1095 Willowdale Road (MS 2027), Morgantown, West Virginia 26505 (E-mail: owirth@cdc.gov).

that automatically detected vocal responses (up, down, left, and right) to study the relation between these vocal responses and corresponding joystick responses. Their results indicated that speech recognition could be used reliably to study vocal operants, allowing precise comparisons to a manual response. As remarkable as this system was for the technology available at the time, it was limited to a few forms of responding and required a combination of computer-controlled speech recognition and voice-activated relays to record the responses and provide differential consequences.

More recently, Manabe, Kawashima, and Staddon (1995) used an elaborate signal-processing apparatus that allowed accurate discrimination of high- and low-frequency vocalizations by budgerigars, and even made possible the differential reinforcement of vocalizations. Although this method was successful for its intended purpose, the complexity of the apparatus and its inability to detect fine differences in the frequencies of vocalizations hinder its broader application.

Recent advances in audio-signal processing and analysis have led to the development of increasingly sophisticated and accurate speech-recognition systems. Today, the limits of speech recognition are related mainly to the speed with which a computer system can respond to an utterance. Along with more efficient speech-recognition algorithms, the fast processors of modern computers have increased recognition accuracy and speed, extended vocabulary size, and minimized training requirements. Given these recent developments in speech-recognition technology and increased affordability of high-speed computers, the application of speech-recognition technology to a broad range of human behavioral research is now feasible.

The purpose of the current study was technological: to illustrate the potential of speech-recognition technology for addressing issues that concern behavioral researchers. We describe how speech-recognition technology can be applied to study a kind of stimulus class formation, functional equivalence, that might be important for understanding aspects of verbal behavior. We also suggest several other areas of research that might benefit from using vocal behavior as its primary datum. We also describe the current state of speech-recognition software and offer some practical suggestions for behavior analysts who are interested in exploring this evolving technology.

METHOD

Participants

Eight undergraduate college students (7 females and 1 male, ages 18 to 29 years) participated. They were recruited through advertisements placed on recruitment bulletin boards located in the Department of Psychology at West Virginia University. Students had no prior experience with similar experimental protocols. All were required to sign an informed consent agreement that described the general procedures. Students were paid 1¢ for each correct response in a block of trials in which an accuracy criterion (approximately 90%) was met. If the accuracy criterion was not met, no earnings were available for that block. In addition, students received a $1 bonus per session for attending all scheduled sessions. Sessions were conducted 3 to 5 days per week and lasted approximately 50 min.

Apparatus and Stimuli

Daily sessions were conducted in a room (2.2 m by 1.8 m) equipped with a table, a chair, and the experimental apparatus. The apparatus consisted of a microcomputer equipped with an 33 MHz 486 processor, 16 MB RAM, an IBM® M-ACPA sound card, a VGA color monitor, headphones, a VXI® headset microphone, and a keyboard. Experimental events and data collection were controlled by C programming. Throughout experimental sessions, each student wore a headset microphone and headphones for auditory feedback and to help mask extraneous noises.

The speech-recognition software used was Dragon System's Dragon VoiceTools™ Version 1.01. This software can be programmed to accompany any C or C++ computer program such that function calls can be made to a memory-recognition speech driver to process speech input. (Programming routines are available from the first author.) When a speaker emits a vocal utterance 25 dB above ambient noise levels, an analog signal from the microphone is converted to digital format in the audio board, and a digital representation of the vocal utterance is sent to the memory-resident speech driver. The digitized pattern then is compared with word patterns stored in memory from a specified vocabulary of utterances sampled from the speaker. The digitized patterns are sent to the speech driver during an utterance so that processing and recognition can occur simultaneously, without waiting for the end of an utterance. The speech driver requires a minimum of 100 ms between utterances. This procedure allows speech input to be handled similarly to keyboard input without appreciably slowing the main computer program. In addition, the speech-recognition driver allows measurement of vocal utterance duration (in milliseconds) and amplitude (in decibels), as well as providing a confidence level that represents the degree to which a spoken word matches the digitized pattern of that word.

Stimuli consisted of white symbols measuring approximately 2 cm by 2 cm on a computer screen (19 cm by 24 cm). Figure 1 shows the two sets of stimuli. On each trial, one stimulus appeared in a blue box (3 cm by 3 cm) at the center of the screen. Each stimulus was assigned a number and a letter for descriptive purposes only. Numbers corresponded with the experimenter-defined stimulus classes, and the letters A, B, and C designated stimulus class members. For example, A1, B1, and C1 were designated as the same functional class. Vocal responses used for naming trials also were assigned numbers that designated corresponding classes of stimuli. For example, Response 1 (GOX) was used to establish the functional stimulus class A1B1C1. Students S101, S103, S105, and S108 were trained with Stimulus Set 1 (top panel of Figure 1), whereas Students S102, S104, S106, and S107 were trained with Stimulus Set 2 (middle panel).

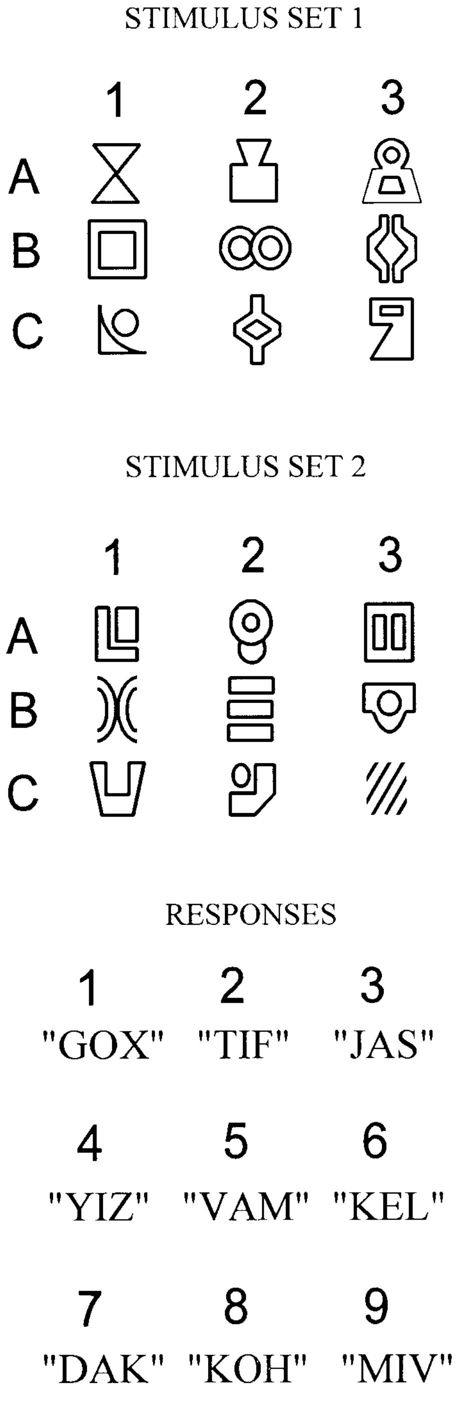

Fig. 1. Sets of stimuli and responses used to establish functional stimulus classes.

Procedure

Speech-recognition training. Training occurred in two stages. First, a sampling of each nonsense word was recorded and saved as a digitized pattern of the utterance. Students were prompted to say out loud each of the three nonsense words (e.g., GOX, TIF, and JAS) five times in succession. The recorded utterances were digitized and the speech pattern adapted and normalized to an individualized ‘‘speech model’’ for that word. At this point in the training, recognition of the nonsense words was fairly accurate; however, a second training phase was conducted to ensure even greater accuracy. In this second training phase, students were prompted to say each nonsense word in random order, one at a time. Each utterance was recorded, digitized, and then compared to the speech models created previously. This comparison yielded software-generated ‘‘confidence’’ values (1 to 100) that indicate the degree to which each utterance matched one or more of the speech models for that student. Utterances meeting a minimum confidence criterion of 40 (a value recommended by Dragon System’s software developers in the software manual) were adapted to existing speech models. Utterances for which the recognition confidence was below criterion were rejected. Training continued until confidence values met or exceeded a value of 90 five times for each nonsense word, regardless of whether this value was met consecutively. The second training phase ensured an adequate sampling of each nonsense word and, as a result, a very high degree of speech-recognition accuracy. Completion of both phases of training lasted approximately 5 min. Our informal evaluations of recognition accuracy yielded no errors across the range of utterances differing in pitch, amplitude, and duration. Periodic checks of speech-recognition accuracy revealed that high accuracy was maintained even after several weeks.

Naming procedure. Blocks of naming trials began with the following instructions displayed on the computer screen:

> During the next set of activities, your job will be to correctly name the symbols. Each trial will begin with the presentation of a symbol positioned inside a blue box at the center of the screen. If you know the correct name, say it out loud. If you don’t know the correct name, wait and the correct name will appear on the screen—say it then. You can only earn money, however, if you make a correct response before it is displayed on the screen. Press ‘‘S’’ when you are ready to start.

Each naming trial began with the presentation of a sample stimulus. If no vocal response occurred within 20 s, students were prompted to make the correct response with the following message: ‘‘Please say the correct name.’’ A correct response produced the word ‘‘Correct’’ at the bottom of the screen for 1 s along with a 50-ms 2000-MHz tone. An incorrect response produced the word ‘‘Incorrect’’ at the bottom of the screen for 1 s and a 50-ms 500-MHz tone. Responses that were not among the set of experimenter-defined responses were not considered either correct or incorrect. Instead, those responses were considered analogous to ‘‘off-key’’ presses that are possible during common conditional discrimination tasks. Following such responses, students were prompted to make another response with the following message displayed on the screen: ‘‘Not recognized—try again.’’ A response was designated as off key when recognition confidence did not meet a minimum criterion of 40. In the present study, off-key responses rarely occurred.

After a recognized response occurred, the screen was cleared except for an empty blue stimulus box, and a variable 0-s to 2-s intertrial interval (ITI) was initiated. If a vocal response occurred during the ITI, a 5-s delay to the presentation of the next sample stimulus resulted. Other responses during naming trials, including pressing the keys, had no programmed consequences. Stimuli were presented in a quasirandom sequence, with the restriction that no stimulus appeared on more than three consecutive trials. Students’ earnings and percentage correct were displayed on the screen following the completion of each trial block, except on test trials when performance feedback was precluded.

Baseline training. A graded delayed-prompt procedure was used to minimize the frequency of errors during initial training of new name relations. The delay between the presentation of a sample stimulus and the display of a written prompt for the correct (class-consistent) response (e.g., ‘‘say JAS’’) was increased gradually from 2 s to 5 s and then was eliminated. Initial trial blocks began with a 2-s delay between the onset of a stimulus and the presentation of the response prompt positioned approximately 2 cm below the stimulus. When performance reached an accuracy criterion (22 of 24 trials correct for three consecutive trial blocks), the delay between the presentation of a stimulus and the correct response was increased to 5 s until performance again reached criterion. There-after, performance was assessed in the absence of response prompts.

Table 1

Composition of trail blocks during baseline training and tests of functional equivalence.

Phase	Trials of each type/ total per block	Trial types (stimulus → response)
Baseline training		
A-R	8/24	A1 → 1 (GOX) A2 → 2 (TIF) A3 → 3 (JAS)
B-R	8/24	B1 → 1 (GOX) B2 → 2 (TIF) B3 → 3 (JAS)
C-R	8/24	C1 → 1 (GOX) C2 → 2 (TIF) C3 → 3 (JAS)
A-R, B-R, C-R mix	4/36	A1 → 1 (GOX) A2 → 2 (TIF) A3 → 3 (JAS) B1 → 1 (GOX) B2 → 2 (TIF) B3 → 3 (JAS) C1 → 1 (GOX) C2 → 2 (TIF) C3 → 3 (JAS)
Test of functional equivalence		
Function Change 1	8/24	A1 → 4 (YIZ) A2 → 5 (VAM) A3 → 6 (KEL)
Transfer Test 1	4/36	A1 → 4 (YIZ) A2 → 5 (VAM) A3 → 6 (KEL) B1 → ? B2 → ? B3 → ? C1 → ? C2 → ? C3 → ?
Function Change 2	4/36	A1 → 7 (DAK) A2 → 8 (KOH) A3 → 9 (MIV)
Transfer Test 2	4/36	A1 → 4 (DAK) A2 → 5 (KOH) A3 → 6 (MIV) B1 → ? B2 → ? B3 → ? C1 → ? C2 → ? C3 → ?

Using the graded delayed-prompt procedure, original baseline name relations were taught in four stages (see baseline training in Table 1). Initial trial blocks consisted of the three A-R trial types (A1-1, A2-2, and A3-3), each presented eight times per block. Upon reaching the accuracy criterion of 22 of 24 trials correct for three consecutive trial blocks, the training of B-R trial types (B1-1, B2-2, and B3-3) and then C-R trial types (C1-1, C2-2, and C3-3) proceeded in the same manner. In the fourth stage of training, a mix of A-R, B-R, and C-R trial types was presented four times in each block, and until the standard accuracy criterion was met. Students then were required to demonstrate accurate performance (i.e., 22 of 24 trials correct) for at least one trial block at each of four levels of reduced feedback (e.g., 75%, 50%, 25%, and 0% of trials).

Tests for functional equivalence. Interchangeability of stimulus functions was used to demonstrate the establishment of functional equivalence among the class members (Goldiamond, 1962, 1966). This was accomplished by training a new response to one member of each class and then testing the remaining class members for a corresponding change in responding (i.e., transfer of function).

Following the establishment of original baseline name relations in baseline training, the new Responses 4, 5, and 6 (YIZ, VAM, and KEL) were introduced to the students' vocabulary. Students' speech models were adapted with new responses by repeating the speech-recognition training procedure with the inclusion of Responses 4, 5, and 6. Following speech-recognition training, speech models now consisted of Responses 1 through 6, and any of these responses were acceptable on subsequent naming trials.

Following speech-recognition training, the new Responses 4, 5, and 6 were reinforced in the presence of Stimuli A1, A2, and A3, respectively, using the graded delayed-prompt procedure. Trial types (e.g., A1-4, A2-5, and A3-6) were presented eight times each for a total of 24 trials per block. Upon reaching the accuracy criterion of 22 of 24 trials correct for each of four levels of reduced feedback (e.g., 75%, 50%, 25%, and 0% of trials), test trials consisting of the A stimuli and each of six B and C stimuli (e.g., B1, B2, B3, C1, C2, and C3) were presented to test for transfer of function. Each stimulus was presented four times per block for a total of 36 trials. Test blocks were conducted in the absence of performance feedback.

Table 2

Number of trial blocks in each phase of baseline training and tests of functional equivalence.

	Students							
Phase	S101	S102	S103	S104	S105	S106	S107	S108
Baseline training								
A-R	11	9	10	10	10	10	9	13
B-R	10	9	9	9	10	10	10	10
C-R	10	9	10	10	10	9	10	10
A-R, B-R, C-R mix	7	7	7	7	7	7	7	8
Tests of functional equivalence								
Function Change 1	14	14	14	14	14	15	14	14
Transfer Test 1	3	3	3	1	1	3	2	3
(criterion met?)	(no)	(no)	(no)	(no)	(no)	(no)	(no)	(yes)
Function Change 2	13	14	14	15	14	13	14	
Transfer Test 2	3	1	2	2	1	2	3	
(criterion met?)	(no)	(yes)	(yes)	(yes)	(yes)	(yes)	(no)	

If transfer of function was not demonstrated, responses that were consistent with transfer were reinforced until accuracy met or exceeded 90% accuracy for one trial block. This procedure was planned because we did not expect the stimuli to become part of a functional class after one iteration of the function-change procedure. In other research, either repeated reversals have been necessary (e.g., Sidman, Wynne, Maguire, & Barnes, 1989; Vaughan, 1988) or subjects have had an experimental history of matching to sample or other class formation procedures (Layng & Chase, in press). The function-change procedure then was repeated with three new Responses 7, 8, and 9 (DAK, KOH, and MIV) that were reinforced in the presence of A stimuli in the same manner described above. When performance met the accuracy criterion, test blocks that consisted of all A, B, and C stimuli were presented again to test for transfer of function.

RESULTS

Table 2 shows the number of trial blocks required to reach criterion in each phase. Naming performances met the accuracy criterion after 9 to 13 trial blocks with each trial type and seven to eight mixed blocks. In general, the training procedure and stringent criteria resulted in extensive practice and few errors with each stimulus. When naming of A1, A2, and A3 was altered in the first function-change phase from Responses 1, 2, and 3 to 4, 5, and 6, respectively, performances met criterion after 14 to 15 trial blocks. As the subsequent tests for a transfer of function revealed, 7 of 8 students required a second iteration of the function-change procedure for transfer of function to B and C stimuli to occur. In other words, even though these students successfully altered their naming of A stimuli using Responses 4, 5, and 6 (YIZ, VAM, and KEL), they continued to name B and C stimuli with the original Responses 1, 2, and 3 (GOX, TIF, and JAS). These students required an additional function-change phase followed by a second series of transfer tests. The new Responses 7, 8, and 9 were trained to A1, A2, and A3, respectively, and then transfer of the new responses to B and C stimuli was tested again. During this second transfer test, transfer of function was demonstrated in 5 of 7 students.

Figure 2 shows the proportion of responses for Student S103 given each sample stimulus. These data were chosen for illustration to compare unsuccessful and successful transfer of function performances. Differentiation of responding occurred rapidly and with minimal errors during A-R, B-R, and C-R phases of training. The classes of stimuli A1B1C1, A2B2C2, and A3B3C3 soon came to control Responses 1, 2, and 3 (GOX, TIF, and JAS), respectively. Differentiation was evident even in the first trial block after new trial types were introduced and was maintained throughout the mixed trial blocks. The delayed-prompt training procedure likely contributed to the rapid acquisition, which was apparent with other students as well.

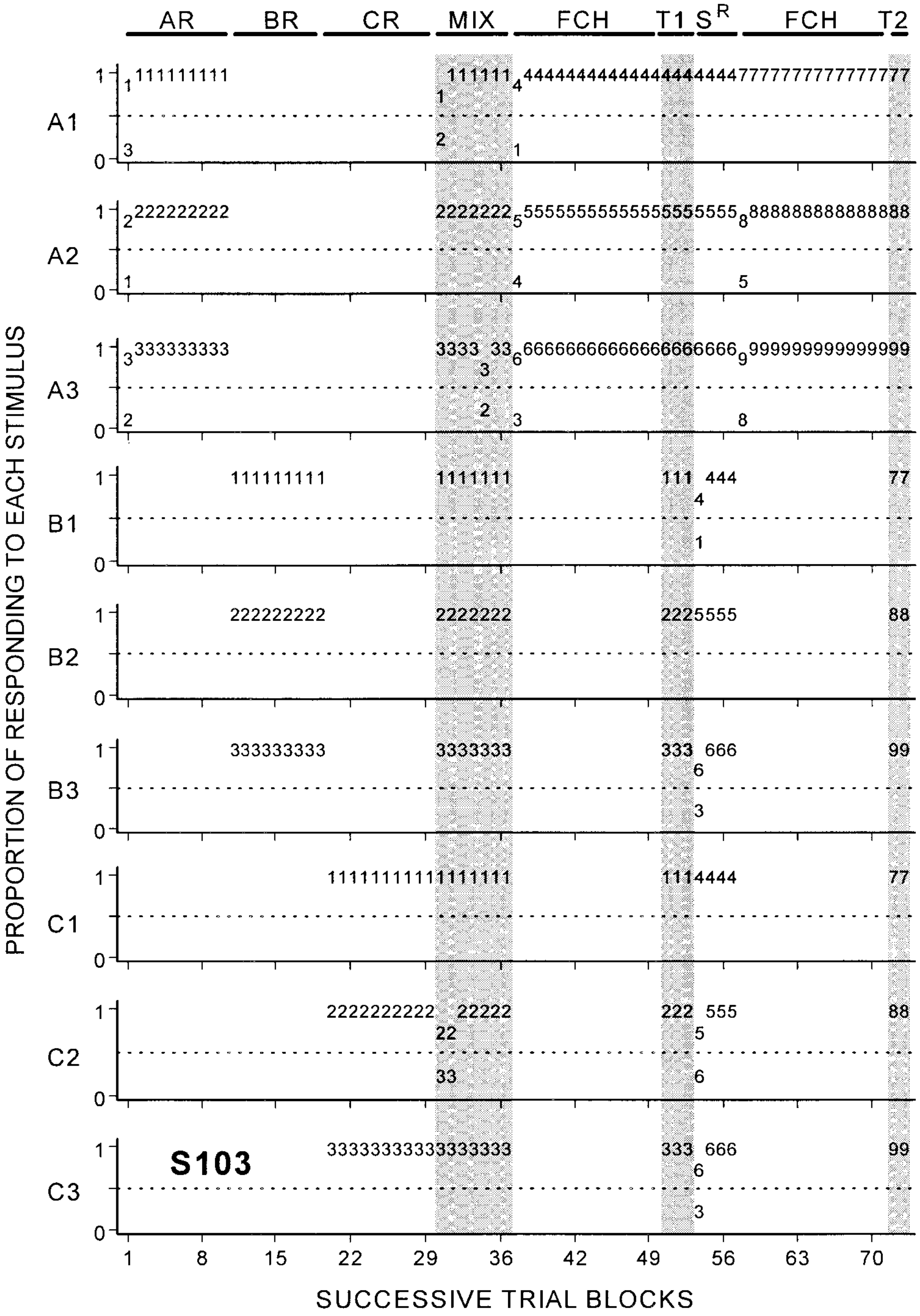

Fig. 2. Proportions of each response type (depicted by number, 1, 2, 3, etc.) that occurred in the presence of each stimulus for Student S103. Proportions are shown across successive trial blocks of each phase of baseline training (A-R, B-R, C-R, and MIX), function changes (FCH), and tests for equivalence (T1 and T2). Shaded portions highlight mixed-trial blocks of A-R, B-R, and C-R trial types during baseline training and tests for functional equivalence. Note that class-consistent responding with Responses 4, 5, and 6 was reinforced (S^R) prior to the second function-change phase.

The function changes involving the A stimuli also occurred rapidly. After some initial disruption in the naming performances, Stimuli A1, A2, and A3 soon came to control the new Responses 4, 5, and 6 (YIZ, VAM, and KEL), respectively. When transfer of Responses 4, 5, and 6 was tested with B and C stimuli, shown under the T1 phase in Figure 2, transfer did not occur. This student, like most students, continued to name B and C stimuli with the original Responses 1, 2, and 3 (GOX, TIF, and JAS). When reinforcement was dependent on these responses in the presence of B and C stimuli, this student's responses changed (see the S^R phase in Figure 2). When the functions of Stimuli A1, A2, and A3 were changed again, the student responded consistently with new Responses 7, 8, and 9 (DAK, KOH, and MIV) (see the second FCH phase in Figure 2). The final test for a transfer of function, T2, shows that the B and C stimuli also came to control the new Responses 7, 8, and 9, suggesting that distinct functional equivalence classes were established.

Other response parameters, such as sample–response speeds (inverse latency in seconds), response duration, and software-generated confidence values, also were recorded. Amplitude of response types in decibels was recorded but is not presented here, because this measure was found to be invariant across all response types and phases of the study. Figure 3 shows the median sample–response speeds recorded for the same student (S103) across training and test phases. In general, sample–response speeds increased with increased exposure to training trials, both during initial baseline training and during function-change phases. Speeds also were somewhat slower during the transfer tests than during the previous training blocks.

Although the speech-recognition software used in the present study was intended for recognition of discrete utterances (i.e., only a single utterance or word was recognized at a time), speeds higher than one per second were recorded for many utterances. This suggests that speech recognition was not only accurate but also fast. In fact, pilot data from our laboratory have shown that recognition rates of 60 utterances per minute are possible when no delays (e.g., ITIs) are programmed between successive response opportunities. The speed of speech-recognition systems opens new avenues of research that requires rapid vocal responding, especially with newer speech-recognition systems (e.g., Dragon NaturallySpeaking® Developer Suite), which are now capable of detecting and keeping pace with continuous speech.

Figure 4 shows utterance duration for each response type across phases for Student S103. Most apparent was that duration corresponded roughly to the topography of the utterance. In other words, some utterances took longer to say than others. For example, Response 6 (KEL) took approximately 175 ms to emit, compared to about 250 ms for Response 5 (VAM) and 400 ms to 500 ms for other responses such as 1, 3, and 4 (GOX, JAS, and YIZ). No systematic differences were found in utterance durations across training and test phases. Despite using one-syllable consonant-vowel-consonant nonsense words in the present study, results showed that utterance duration depended heavily on the phonetic structure of the words. Therefore, if duration of responding is to be taken as a critical measure of responding, it is important to take into account variations in the time it takes to produce an utterance, and to choose words that require similar production times.

Figure 5 shows the software-generated confidence values for each response type across phases of the study for Student S103. It is important to note that accuracy of utterance recognition remained high throughout the experiment. Variations in confidence values merely represented the degree of variation between the speech pattern of a recorded utterance and the model. Variations in response parameters including duration, amplitude, and even specific frequencies may alter confidence values. Other factors may also contribute, such as the number of words in the software's active vocabulary from which a recorded utterance must be discriminated. Confidence values may decrease if utterances have to be compared to many speech models in the active vocabulary. Furthermore, the speech model for each word was adapted continually with each utterance to represent slight variations in the way a word is spoken.

Confidence values for Responses 1, 2, and 3 (GOX, TIF, and JAS) decreased somewhat from initial training to the mixed training blocks. Confidence values for these same re-

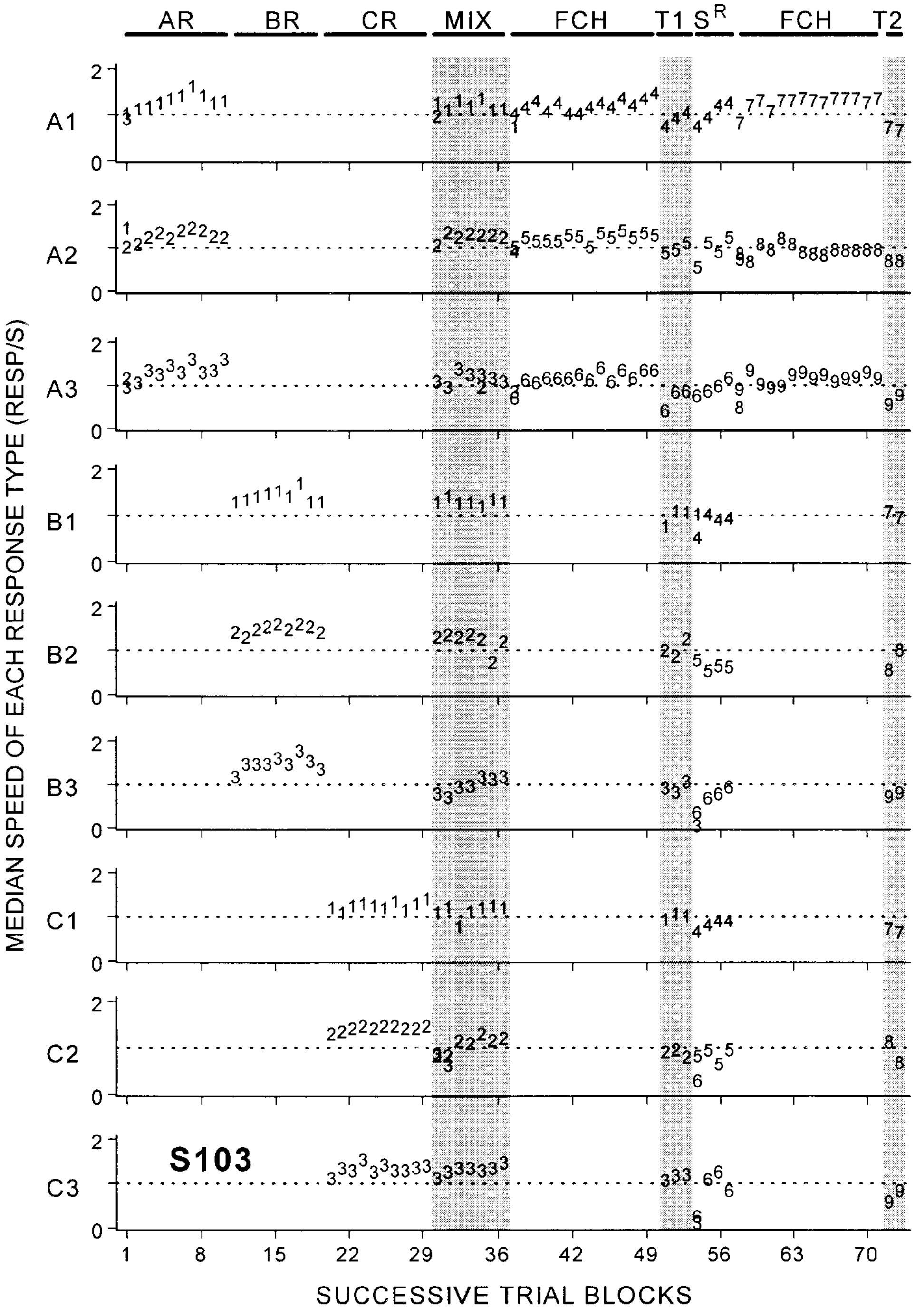

Fig. 3. Median sample–response speed (1/latency) of each response type (1, 2, 3, etc.) in the presence of each stimulus for Student S103. Speeds are shown across successive trial blocks of each phase of baseline training (A-R, B-R, C-R, and MIX), function changes (FCH), and tests for equivalence (T1 and T2). All other details as in Figure 2.

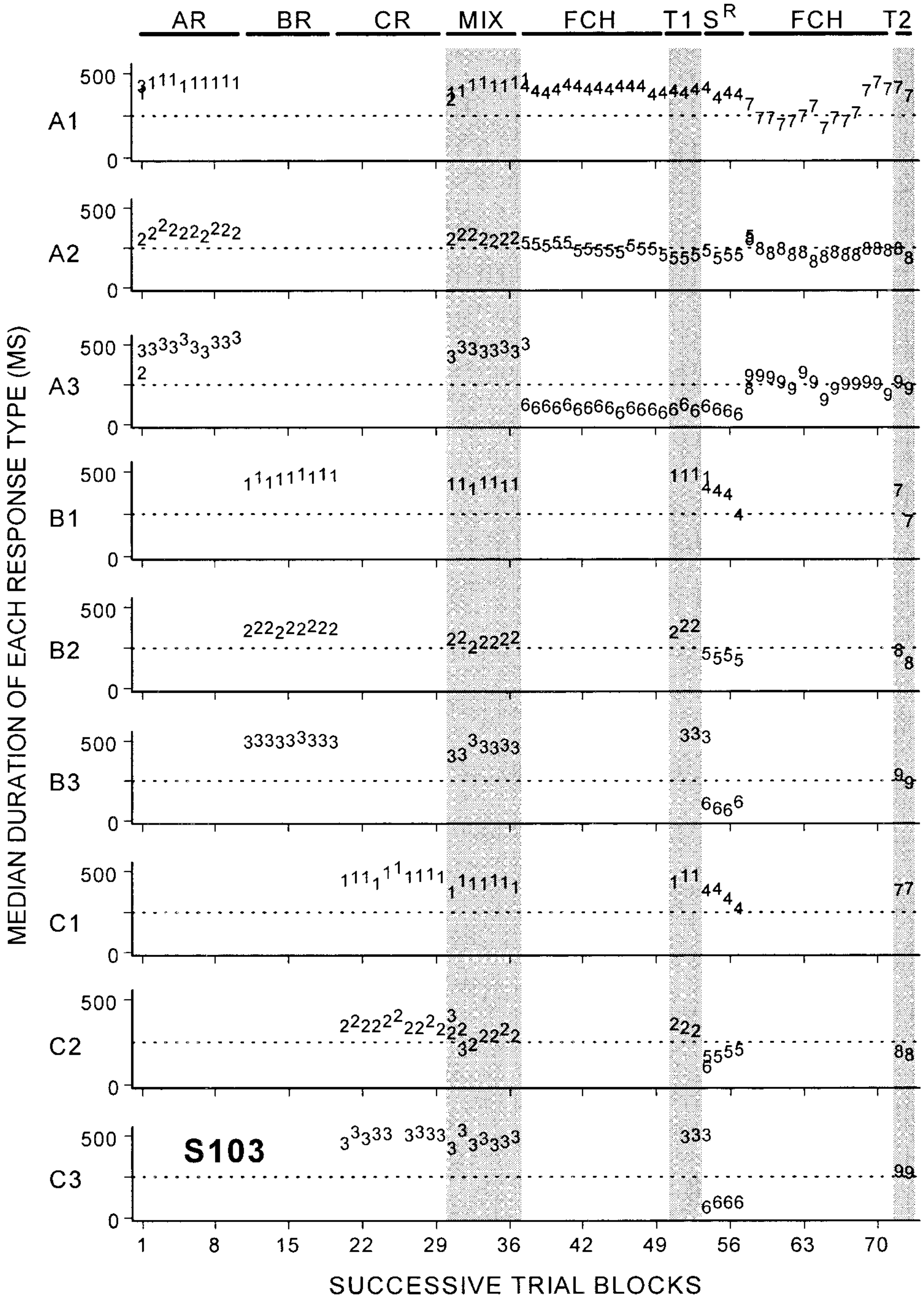

Fig. 4. Median duration of each response type (1, 2, 3, etc.) in the presence of each stimulus for Student S103. Durations are shown across successive trial blocks of each phase of baseline training (A-R, B-R, C-R, and MIX), function changes (FCH), and tests for equivalence (T1 and T2). All other details as in Figure 2.

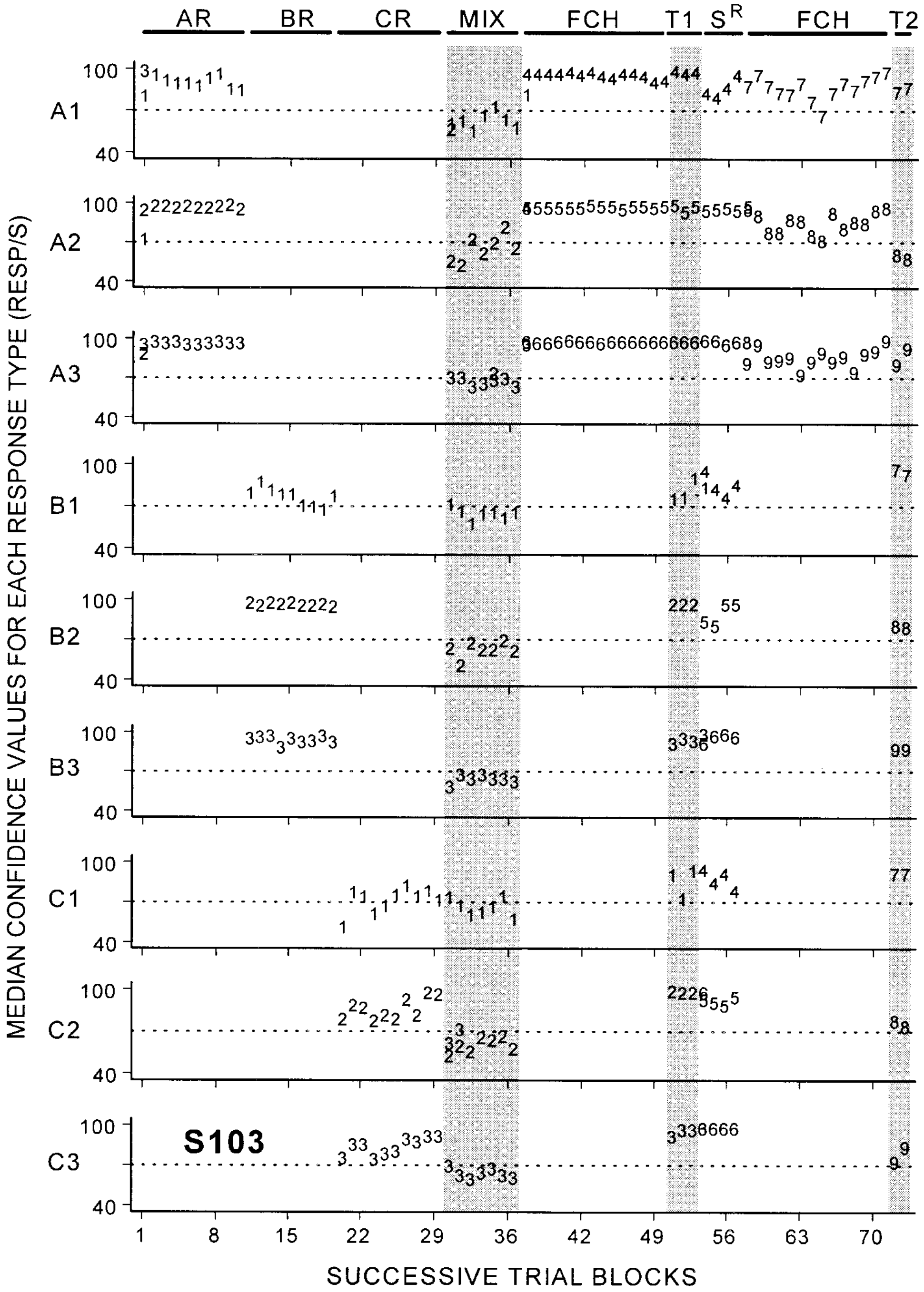

Fig. 5. Median confidence value for each response type (1, 2, 3, etc.) in the presence of each stimulus for Student S103. Confidence values are shown across successive trial blocks of each phase of baseline training (A-R, B-R, C-R, and MIX), function changes (FCH), and tests for equivalence (T1 and T2). All other details as in Figure 2.

sponses later increased during the first transfer test (T1). The increase and the relatively high confidence values for Responses 1 through 6 were most likely the result of repeating the speech-recognition training just prior to the first function-change phase. During speech-recognition retraining, the original Responses 1, 2, and 3 were retrained along with new Responses 4, 5, and 6, which were introduced and trained for the first time. Retraining, as programmed in the present experiment, recreated the speech models for the original Responses 1, 2, and 3 from scratch, thus removing accumulated variations in speech pattern that had been adapted to the models over time.

Systematic differences in confidence values also occurred across different words. For example, Figure 5 shows that confidence values for Responses 7, 8, and 9 (DAK, KOH, and MIV) were more variable than those for Responses 4, 5, and 6 (YIZ, VAM, and KEL). Like differences recorded in utterance duration, the phonetic characteristics of these words may have contributed to the software's level of confidence in discriminating utterances.

DISCUSSION

The present results illustrate the use of speech-recognition technology to study functional equivalence, and extend the demonstration of functional equivalence among stimuli to conditions in which shared vocal responses define the stimulus classes. The results showed that utterances were discriminated accurately, and subtle changes in stimulus functions were tracked during transition states. The use of speech-recognition technology not only makes possible an analysis of stimulus control relations involved in naming but also ensures accurate recording and analysis of all vocal responses.

Speech-recognition technology has advanced rapidly in recent years. Several speech-recognition software packages designed for use by novice computer users are now available. Currently, popular and inexpensive versions of speech-recognition software are being shipped along with popular word processing software (e.g., Corel WordPerfect®).

Most of these software packages, however, are intended for end users. Researchers interested in incorporating speech recognition into experimental protocols will require packages designed for software developers. At present, we know of only a few commercial products designed specifically for software developers: Dragon System's Naturally-Speaking® software development kit (SDK), Lernout and Hauspie's Voice XPress™ SDK, Microsoft's Whisper Speech Recognizer SDK, and IBM's ViaVoice™ SDK. The version of the software by Dragon Systems that was used in the present research has been replaced with a Microsoft Windows® version that incorporates newer speech-processing algorithms that make continuous-speech recognition possible (i.e., no pauses are required between utterances). The SDK software products that are currently available typically make use of ActiveX components that can be easily incorporated into user-developed software programs using development environments such as Microsoft Visual Basic or Visual C++. A modicum of programming skill appears to be sufficient to incorporate simple speech-recognition capability into one's research protocol. An intermediate to advanced level of computer programming skill is recommended to extract more advanced measures (e.g., to extract the raw waveform of vocal utterances, calculate the amplitude, or generate a frequency spectrograph).

Other emerging technologies may further advance the sophistication of speech-recognition technology and encourage its use in experimental research. The technology domain of ''interactive voice response'' is being developed by leading software manufacturers. For example, a software manufacturer's consortium that includes AT&T, IBM, Lucent, and Motorola has developed Voice eXtensible Markup Language (VoiceXML). VoiceXML is an XML-based markup language that can be used for distributed (i.e., networked) voice applications, much as HTML is a language for distributed visual applications. VoiceXML is designed for applications that feature synthesized speech, digitized audio, and recognition or recording of spoken input. A similar technology that sets standards for distributed voice applications is available from the Microsoft Corporation (i.e., the Speech Application Programming Interface; SAPI). These standards-based programs are likely to produce a reliable and ubiquitous speech-recognition technology that may eventually replace other types of computer interface elements (e.g., keyboards).

In conclusion, productive science is marked by the development of effective measurement. Speech-recognition technology has advanced sufficiently to provide measures of vocal responding that meet this criterion. Many examples of interesting verbal behavior discussed by Skinner (1957) and others might be facilitated through the use of speech-recognition technology. Some of these include the automated recording and transcribing of vocal responding that could facilitate the analysis of the role of naming in verbal behavior and in the formation of stimulus classes (e.g., Horne & Lowe, 1996), problem solving (e.g., Ericsson & Simon, 1984; Hayes, 1986), verbal self-reports (e.g., S. D. Lane & Critchfield, 1996), and even nonhuman vocalizations (e.g., Manabe et al., 1995). Speech recognition may be just the measure the field has been waiting for to further advance the science of verbal behavior.

REFERENCES

Baron, A., & Journey, J. (1989). Reinforcement of human reaction time: Manual-vocal differences. *The Psychological Record, 39,* 285–296.

Cross, D., & Lane, H. (1962). On the discriminative control of concurrent responses: The relations among response frequency, latency, and topography in auditory generalization. *Journal of the Experimental Analysis of Behavior, 5,* 487–496.

Ericsson, A. K., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Eshleman, J. (1991). Quantified trends in the history of verbal behavior research. *The Analysis of Verbal Behavior, 9,* 61–80.

Flanagan, B., Goldiamond, I., & Azrin, N. (1958). Operant stuttering: The control of stuttering behavior through response-contingent consequences. *Journal of the Experimental Analysis of Behavior, 1,* 173–177.

Goldiamond, I. (1962). Perception. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology* (pp. 280–340). New York: Basic Books.

Goldiamond, I. (1966). Perception, language, and conceptualization rules. In B. Kleinmuntz (Ed.), *Problem solving* (pp. 183–224). New York: Wiley.

Hake, D. F., & Mabry, J. (1979). Operant and nonoperant vocal responding in the mynah: Complex schedule control and deprivation-induced responding. *Journal of the Experimental Analysis of Behavior, 32,* 305–321.

Hayes, S. C. (1986). The case of the silent dog—Verbal reports and the analysis of rules: A review of Ericsson and Simon's *Protocol Analysis: Verbal Reports as Data. Journal of the Experimental Analysis of Behavior, 45,* 351–363.

Horne, P., & Lowe, C. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185–241.

Lane, H. (1960). Temporal and intensive properties of human vocal responding under a schedule of reinforcement. *Journal of the Experimental Analysis of Behavior, 3,* 183–192.

Lane, H. (1964). Differential reinforcement of vocal duration. *Journal of the Experimental Analysis of Behavior, 7,* 107–115.

Lane, H., & Shinkman, P. (1963). Methods and findings in an analysis of vocal operant. *Journal of the Experimental Analysis of Behavior, 6,* 179–188.

Lane, S. D., & Critchfield, T. S. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior, 65,* 355–374.

Layng, M. P., & Chase, P. N. (in press). Stimulus-stimulus pairing, matching-to-sample testing, and emergent relations. *The Psychological Record.*

Leander, J., Milan, M., Jasper, K., & Heaton, K. (1972). Schedule control of the vocal behavior of *Cebus* monkeys. *Journal of the Experimental Analysis of Behavior, 17,* 229–235.

Manabe, K., Kawashima, T., & Staddon, J. (1995). Differential vocalization in budgerigars: Towards an experimental analysis of naming. *Journal of the Experimental Analysis of Behavior, 63,* 111–126.

Milheim, W. (1993). Computer-based voice recognition: Characteristics, applications, and guidelines for use. *Performance Improvement Quarterly, 6,* 14–25.

Miller, L. (1968). Escape from an effortful situation. *Journal of Applied Behavior Analysis, 11,* 619–627.

Molliver, M. (1963). Operant control of vocal behavior in the cat. *Journal of the Experimental Analysis of Behavior, 6,* 197–202.

Routh, D. (1969). Conditioning of vocal response differentiation in infants. *Developmental Psychology, 1(3),* 219–226.

Salzinger, K., Waller, M., & Jackson, R. (1962). The operant control of vocalization in the dog. *Journal of the Experimental Analysis of Behavior, 5,* 383–389.

Shearn, D., Sprague, R., & Rosenzweig, S. (1961). A method for the analysis and control of speech rate. *Journal of the Experimental Analysis of Behavior, 4,* 197–201.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Tucker, P., & Jones, D. (1991). Voice as an interface: An overview. *International Journal of Human-Computer Interaction, 3,* 145–170.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 36–42.

Received January 17, 2000
Final acceptance September 25, 2000